ViCoCoS-3D: Videoconferencing common scenes

Conference Paper · June 2016

CITATIONS
0

READS
27

5 authors, including:

Jesus Gutierrez
University of Nantes
19 PUBLICATIONS **107** CITATIONS
SEE PROFILE

César Díaz
Universidad Politécnica de Madrid
10 PUBLICATIONS **15** CITATIONS
SEE PROFILE

Pablo Carballeira
Universidad Politécnica de Madrid
21 PUBLICATIONS **31** CITATIONS
SEE PROFILE

ViCoCoS-3D: Videoconferencing common scenes

Jesús Gutiérrez, César Díaz, Pablo Carballeira, Fernando Jaureguizar, and Narciso García
Grupo de Tratamiento de Imágenes, ETSI Telecomunicación, Universidad Politécnica de Madrid, Spain

***Abstract*—This paper presents a 3D video dataset containing sequences with typical content from videoconferencing scenarios. The objective of this dataset is to provide freely-available sequences for the research community to support the development and evaluation of processing techniques applicable to 3D videoconferencing systems. Therefore, a detailed description of the generation process and the content characteristics is provided, together with insights of possible applications of the dataset.**

***Index Terms*—3D video, dataset, videoconferencing, quality of experience**

I. Introduction

Videoconferencing is nowadays a common alternative to telephony communications and face-to-face meetings both in personal and professional environments, providing also interesting applications like multiparty telemeetings and remote learning. However, there is still a need of improving these systems to provide more realistic, interactive, and immersive audiovisual experiences to the users. Actually, in addition to several recent research works [1][2][3], the importance of this issue is reflected by the activities carried out by the ITU-T study groups 12 (question 10) and 9 (question 12).

One option under research to improve the Quality of Experience (QoE) of the end users of these services is based on using 3D video technologies, founded on the belief in their possibilities to provide an added value to conventional videoconferencing services [1]. Although nowadays there is some skepticism about the success of current 3D video technologies in the entertainment consumer market, videoconferencing, together with medical applications and gaming, seems to be one of the domains where 3D video may be more useful [4]. This is mainly motivated by the development of autostereoscopic displays and their arrival to the consumer market, especially in personal devices, such as tablets and smartphones.

In this regard, some studies have been presented analyzing the performance of 3D videoconferencing systems pointing out that there is still room for improving the QoE of the users compared to conventional services [1][2]. Thus, more research is still needed in that respect to support the development of high-quality systems [5]. In this sense, not only the evaluation of real systems is required, but also the performance of processing techniques (e.g., encoding, transmission, etc.) applicable to these services, which entails the necessity of audiovisual content simulating typical videoconferencing scenarios [3].

Therefore, the availability of public video databases is a major requirement for these purposes, allowing the access to useful content for developing and evaluating video processing algorithms and systems, together with the possibility of comparing the results of different approaches and studies using the same test data. In this sense, an extensive work has been carried out by Qualinet to generate a compilation of datasets for several research purposes in relation with image and video processing, which can be found in [6]. However, in contrast to the large amount of conventional video databases [7], only a few include 3D sequences [8], and there is a notable lack of content related to videoconferencing scenarios.

Taking this into account, this paper presents a freely-available dataset of stereoscopic videos, called ViCoCoS-3D (VideoConference Common Scenes in 3D), with typical content from videoconferencing scenarios, aiming at helping in the research on developing and evaluating the performance of processing algorithms and systems for these services. The dataset is available in our website [9].

II. The ViCoCoS-3D dataset

A. Content generation

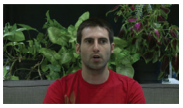
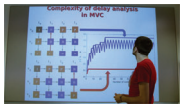
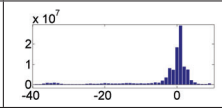
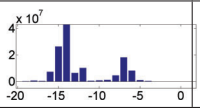
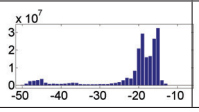
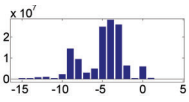
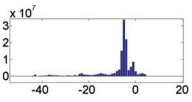
The acquisition of the stereo sequences was carried out using a professional dual-lenses camcorder for broadcasting Sony PMW-TD300 [10]. In particular, the sequences were captured with Full-HD resolution per view and with a framerate of 25 fps and stored in SxS cards of the camera. This storage entails a high-quality compression of the video using MPEG-2, main profile, high-level, YUV 4:2:0 color sampling, and an average bitrate of 35 Mbps. The stored videos for each view were post-processed with *Mistika 2K* to guarantee that the sequences are distortion free, and converted into uncompressed (yuv420p) videos at 25 fps using *FFmpeg*.

In addition, depth maps were computed for each content from both stereo views using the algorithm proposed by Werlberger *et al*. [11]. Although the properties of these depth maps (obtained just from a stereo pair) may limit their applications, they could be useful for developing perceptual models for processing techniques for the considered scenarios.

B. Characteristics of the sequences

The main characteristics of the sequences of the database are shown in Table I. In addition to a description of the scene and the duration, information about the spatial and temporal complexity of the sequences is provided according to the spatial and temporal perceptual information (SI and TI) defined in the recommendation ITU-T P.910 [12]. Furthermore, these indicators were also computed over the depth maps obtained from the stereo sequences to provide information about the spatial (DSI) and temporal (DTI) complexity of the depth content. Finally, previews of the first frame of the videos (left view) and the depth maps are depicted, as well as disparity histograms. Currently, the five sequences are provided without

TABLE I
PROPERTIES OF THE SEQUENCES

Sequence	Screen	Videochat	Meeting	Living_room	Lecture
Description	Man speaking in front of a screen.	Typical videochat scene.	Multiparty business telemeeting.	Multiparty videochat in home environment.	Remote learning scenario.
Frames	400	500	750	500	500
SI	110.06	39.61	53.21	58.45	66.18
TI	18.04	4.42	9.02	6.42	10.04
DSI	28.78	19.10	23.98	22.99	23.90
DTI	5.59	4.13	8.32	3.74	13.50
Preview					
Depth Map					
Disparity Histogram					

audio, since they are addressed to study visual issues related to 3D videoconferencing.

III. APPLICATIONS

The primary objective for the creation of ViCoCoS-3D database was to have realistic content from videoconferencing scenarios available for developing robust systems for delivering this type of services. Specifically, in videoconferencing applications, the attention of the users is focused on very specific parts of the picture, which motivates the use of coding and transmission techniques based on perceptual models related to the 3D visual experience of the users. For instance, error resilience and protection strategies (e.g., FEC protection, traffic shaping, etc.) could be used to reduce the especially annoying impact of transmission errors in the QoE, by means of prioritizing the most perceptually important areas [13]. Similarly, coding techniques could be implemented taking into account perceptual factors related to stereoscopic vision (e.g., binocular suppression or binocular rivalry) in order to provide a high QoE avoiding annoying degradations.

IV. CONCLUSION

This paper presents the ViCoCoS-3D public database providing five high-quality stereoscopic video sequences simulating various videoconferencing scenarios. The main objective of this database is to provide useful content for developing and evaluating video coding and transmission techniques for this type of services. Future work will be focused on extending the dataset including more sequences with realistic properties of videoconferencing applications in terms of audio and video content, and duration to properly evaluate the end users' QoE.

V. ACKNOWLEDGMENT

The authors would like to thank Esther Madejón for her help in the generation of this dataset. This work was supported by the Ministerio de Economía y Competitividad of the Spanish Government under projects TEC2013-48453 (MR-UHDTV) and RTC-2015-4133-7 (MoViSe).

REFERENCES

[1] H. Rizek, K. Brunnström, K. Wang, B. Andrén, and M. Johanson, "Subjective evaluation of a 3D videoconferencing system," in *Proc. of SPIE*, San Francisco, US, Mar. 2014, p. 90110A.

[2] S. Hemami, F. Ciaramello, S. Chen, N. Drenkow, D. Y. Lee, S. Lee, E. Levine, and A. McCann, "Comparing user experiences in 2D and 3D videoconferencing," in *IEEE International Conference on Image Processing*, Orlando, US, Sep. 2012, pp. 1969–1972.

[3] G. Berndtsson, M. Folkesson, and V. Kulyk, "Subjective quality assessment of video conferences and telemeetings," in *International Packet Video Workshop*, Munich, Germany, May 2012, pp. 25–30.

[4] F. Dufaux, B. Pesquet-Popescu, and M. Cagnazzo, *Emerging Technologies for 3D Video*. John Wiley & Sons, Jun. 2013.

[5] J. Skowronek and A. Raake, "Conceptual model of multiparty conferencing and telemeeting quality," in *International Workshop on Quality of Multimedia Experience*, Costa Navarino, Greece, May 2015, pp. 1–6.

[6] Qualinet, "Qualinet Databases." [Online]. Available: http://dbq.multimediatech.cz/

[7] S. Winkler, "Analysis of Public Image and Video Databases for Quality Assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 616–625, Oct. 2012.

[8] M. Urvoy, J. Gutiérrez, M. Barkowsky, R. Cousseau, Y. Koudota, N. García, V. Ricordel, and P. Le Callet, "NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences," in *International Workshop on Quality of Multimedia Experience*, Yarra Valley, Australia, Jul. 2012, pp. 109–114.

[9] Grupo de Tratamiento de Imágenes, Universidad Politécnica de Madrid, "ViCoCoS-3D Dataset," 2016. [Online]. Available: http://www.gti.ssr.upm.es/research/gti-data/databases.html

[10] Sony, "3D Camcorder PMW-TD300." [Online]. Available: http://www.sony.co.uk/pro/product/broadcast-products-camcorders-xdcam/pmw-td300/overview/

[11] M. Werlberger, T. Pock, and H. Bischof, "Motion estimation with non-local total variation regularization," in *IEEE Conf. on Computer Vision and Pattern Recognition*, San Francisco, US, Jun. 2010, pp. 2464–2471.

[12] ITU, "Subjective video quality assessment methods for multimedia applications," *ITU-T Recommendation P.910*, Apr. 2008.

[13] C. Díaz, J. Gutiérrez, J. Cabrera, F. Jaureguizar, and N. García, "Perceptual distortion modeling for side-by-side 3D video delivery," in *IEEE Int. Conf. on Consumer Electronics*, Las Vegas, US, Jan. 2014, pp. 418–419.